

Preuss, Nora; Brändle, Laura S; Hager, Oliver M; Haynes, Melanie; Fischbacher, Urs; Hasler, Gregor (2016). Inconsistency and social decision making in patients with borderline personality disorder. *Psychiatry research*, 243, pp. 115-122. Elsevier

## **Inconsistency and Social Decision Making in Patients with Borderline Personality Disorder**

Nora Preuss<sup>a\*</sup>, Laura S. Brändle<sup>b\*</sup>, Oliver M. Hager<sup>c</sup>, Melanie Haynes<sup>b</sup>, Urs

Fischbacher<sup>d,e</sup>, Gregor Hasler<sup>b\*\*</sup>

<sup>a</sup>Institute of Psychology, University of Bern, 3000 Bern 60, Switzerland

<sup>b</sup>Psychiatric University Hospital, University of Bern, 3000 Bern 60, Switzerland

<sup>c</sup>Laboratory for Social and Neural Systems Research, Department of Economics, University of Zurich, 8006 Zürich

<sup>d</sup>Department of Economics, University of Konstanz, 78457 Konstanz, Germany

<sup>e</sup>Thurgau Institute of Economics, 8280 Kreuzlingen, Switzerland

\*Nora Preuss and Laura S. Brändle contributed equally to the study

**Word counts:** abstract, 150; article body, 4372.

Number: tables, 4; figures, 3; supplemental material, 1

**\*\*Corresponding author:** Gregor Hasler, M.D.

University Hospital of Psychiatry, University of Bern

Bolligenstrasse 111, 3000 Bern, Switzerland.

Tel.: +41 31 930-9543; fax: +41 31 930 99 21.

Email: [gregor.hasler@puk.unibe.ch](mailto:gregor.hasler@puk.unibe.ch), [g.hasler@bluewin.ch](mailto:g.hasler@bluewin.ch)

## Abstract

Inconsistent social behavior is a core psychopathological feature of borderline personality disorder. The goal of the present study was to examine inconsistency in social decision-making using simple economic social experiments. We investigated the decisions of 17 female patients with BPD, 24 patients with major depressive disorder (MDD), and 36 healthy controls in three single shot economic experiments measuring trust, cooperation, and punishment. BPD severity was assessed using the Zanarini Rating Scale for BPD. Investments across identical one-shot trust and punishment games were significantly more inconsistent in BPD patients than in controls. Such inconsistencies were only found in the social risk conditions of the trust and punishment conditions but not in the non-social control conditions. MDD patients did not show such inconsistencies. Furthermore, social support was negatively correlated with inconsistent decision-making in the trust and punishment game, which underscores the clinical relevance of this finding.

Key words: Game theory, borderline personality disorder, behavioral economics, interpersonal behavior, intra-individual variability

## 1. Introduction

Borderline personality disorder (BPD) is a severe and debilitating psychiatric disorder that involves problems in different areas of social and psychological functioning. BPD psychopathology has been described as “stable instability” (Schmideberg, 1959). This instability includes unstable mood and identity, ambivalent relationships, emotional instability, behavioral impulsivity (Doll et al., 2013; Leichsenring et al., 2011), and increased intra-individual variation in neuropsychological tests (Beblo et al., 2006). Taken together, these findings suggest that unstable and inconsistent cognitive and behavioral traits underlie the multifaceted psychopathology of BPD.

Social behavior and compliance with social norms play an important role in human interactions (Spitzer et al., 2007). Considering that problems in interpersonal relations are a core domain within BPD psychopathology (Agrawal et al., 2004), it is appropriate to investigate social behavior. Because of difficulties that arise when quantifying and operationalizing social behavior, such investigations are scarce. Experimental methods based on game theory developed in behavioral economics open up new possibilities of assessing and investigating social preferences such as trust, cooperation, or reaction to punishment through relatively simple social experiments. Such experiments consist of real social interactions involving real monetary incentives (Spitzer et al., 2007) and have been successfully applied to characterize social psychopathology across various psychiatric conditions. Using a game-theoretic approach offers the opportunity to better understand impairments in social interactions in psychopathological conditions such as BPD.

Previous studies that investigated social trust in BPD patients with a game-theoretic approach used repeated interactions with the same partners. In a study by Unoka et al. (2009) participants performed a trust game where an investor had to

interact with a trustee. The investor could share an amount of money with the trustee and the trustee could honor the investor's trust by returning some money. The authors reported reduced trusting behavior in BPD patients (see also Seres et al., 2009). Notably, these results were present only in the trust condition and not in the risk control condition where the investor could transfer money to a lottery. In a further study, King-Casas et al. (2008) were able to show that BPD patients are unable to maintain cooperation in a trust game. When HC played the investor role with a BPD trustee, investment levels were particularly lower in late rounds of a trust game suggesting a break in cooperation caused by social signals emitted by the BPD trustee (King-Casas et al., 2008). BPD participants not only ruptured cooperation, they were also less likely to show generous gestures to repair their break. The authors mentioned that a successful cooperation between two agents requires a range of intact computations (see also Firth & Firth 2006; Singer 2006, Sanfey et al., 2006). Trust and cooperation, however, can easily be ruptured by unpredictable behavior, such as observed in BPD.

An unstable self-concept and unpredictable and inconsistent interpersonal behaviors are key features in BPD (Lieb et al., 2004). BPD symptoms have frequently been attributed to an impairment in the underlying attachment organization (Levy, 2005). According to the attachment theory (Ainsworth et al., 1978; Bowlby, 1969; 1973; 1977; 1980), the affective bond between a child and a caregiver has an influence on the emerging self-concept of the child (for overview see Levy, 2005). Upon the interactions with the caregiver (or attachment figure), children build affective-cognitive schemata of the self and others – so called “internal working models”. These models are thought to influence and organize personality development and to shape future relationships (Bowlby, 1969). Additionally, they are relatively stable across lifespan (Collins & Read, 1994). Fonagy (2000) proposed that

a secure attachment is important because it offers the child the possibility to safely explore the mind of a caregiver and by doing so to learn about minds in general (see also Fonagy et al., 2000). Fonagy et al. (2000) called this ability *mentalization*, which is the ability to understand one's own as well as others' behavior and thoughts. He proposed that reflective and responsive parents increase the likelihood of a secure attachment which in turn positively influences the ability to mentalize. Mentalizing includes the capacity to generalize other's intentions to new social situations and to adjust evaluations in repeated social interactions. Securely attached individuals with BPD are very rare (Levy, 2005). The childhood and youth of BPD patients are typically characterized through conditions that impairs mentalizing (Ball & Links, 2009). Neglect, abuse, maltreatment through caregivers are the most obvious features of such conditions. Even in the absence of such significant adversities, caregivers frequently show severe inconsistencies in their social behavior (see Clarkin et al., 2007). They serve as a source of both fear and reassurance. As a response, their children lose their sense of a stable self and do not develop the capacity to predict other's people thoughts and actions. Their impaired ability to mentalize contributes finally to disorganized attachment and increased risk for BDP. Consistent with this theory, a recent study was able to show that social support serves as protective factor of developing BPD (Rosenbach and Renneberg, 2014). We now propose that disorganized attachment styles in BPD result in inconsistent and unpredictable behavior when it comes to social interaction. BPD patients are unable to anticipate intentions or behavior of others, and therefore in turn exhibit an unpredictable behavior. The main goal of the present study was therefore to examine behavioral inconsistency in BPD using a game-theoretic approach. We used one-shot social exchange games in order to exclude more complex influences of reputation formation and learning from feedback. This enabled us to investigate

social behavior in first encounters. Based on the inability to mentalize about other peoples' thoughts and intentions the adopted relationship heuristic in BPD should also lead to inconsistent behavior in first encounters. We assessed social behavior in various situations and contexts in order to evaluate the specificity of our findings. We conducted a trust game, a punishment game and a coordination game. We hypothesized that behavioral inconsistencies are limited to social contexts and will not be present in non-social control conditions. Based on the finding that social support serves as protective factor for developing BPD, we proposed that behavioral inconsistencies are associated with low perceived social support. Based on the findings of Unoka et al., (2009) we further assumed that BPD patients should show decreased investment in the trust and punishment game. Furthermore, we expected that BPD patients would be less likely to adjust their investment behavior in a cooperation game given a prior interaction with a partner, because they are unable to mentalize about the intentions of the partner. To account for diagnostic specificity, we included two control groups: healthy controls (HC) and patients with major depressive disorder (MDD).

## **2. Methods**

### **2.1 Subjects**

Participants included 17 female BPD patients and 2 control groups: 24 female MDD patients, and 36 female HC. HC were recruited via flyers, newspaper advertisements, and personal contact at the University Hospital of Psychiatry in Bern, where they were assessed. Patients (BPD and MDD) were recruited via personal contact at the University Hospital of Psychiatry in Bern. Psychopathology was assessed using the Structured Clinical Interview (SKID) for DSM-IV axes I (First et al., 2002) and II (First et al., 1997) disorders in both patients and HC control group.

Experienced clinicians and researchers performed diagnostics and symptom rating. Inclusion criteria for HC were no history of psychiatric or physical disorders and no current drug or medication use. The inclusion criterion for BPD patients was a current diagnosis of BPD. The inclusion criterion for MDD patients was a current diagnosis of MDD. Other psychiatric disorders such as affective disorder, eating disorder, or substance abuse were not part of our exclusion criteria as long as BPD/MDD symptoms were the most prevalent. Exclusion criteria were psychotic disorder or drug and alcohol abuse. Specific BPD symptom occurrence during the last week was assessed using the Zanarini Rating Scale for Borderline Personality Disorders ZAN-BPD (Zanarini, 2003), and education level was measured using a rating scale ranging from 0 (no school finished) to 9 (high university degree). Current clinical symptoms were assessed using the Beck Depression Inventory (BDI) (Beck et al., 1961) and the Beck Anxiety Inventory (BAI) (Beck et al., 1988). Education level and body mass index (BMI) were assessed through self-report. To assess experienced social support the 14 item *Questionnaire on Social Support* was used (F-SozU K-14, Fydrich et al., 2009). *The Questionnaire on Social Support* measures how a participant perceives social support (including emotional support, practical support and social integration) from his or her social network. Higher values indicate higher perceived social support. Current medication use was no exclusion criterion in BPD and MDD patients.

Prior to the actual experiment, 64 random participants were recruited that served as interaction partners during the social games. These participants did not undergo any prior clinical assessment.

After describing the study to the subjects, written informed consent was obtained. All aspects of the study were approved by the regional ethics committee

(Kantonale Ethikkommission Bern) and were in accordance with the Declaration of Helsinki principles.

## **2.2 Neuroeconomic games**

To investigate social behavior we used three different economic experiments: the trust game, cooperation game, and punishment game. In each game, player A and player B interacted anonymously. Players A were the patients with BPD, MDD or the group of healthy controls. Players B were a group of randomly selected participants, that were not screened for any pathologies. Players B solely served as interaction partners for Players A and their responses to each possible actions of player A were recorded in a previous session (strategy method). As an example, they were asked how many points they would return if a Player A would give them 0, 10, 20 ... 100 points in the Trust Game Players' B responses were recorded by the experimenter and then used in the actual experiment with player A. Players' B decisions were also payoff relevant because they were later paid based on their own decision and the decisions of players A. All players were informed that they interacted with a human player or a computer. Every decision in the games involved a distribution of monetary units (MUs), which were converted into money at the end of the experiment. Participants were explicitly told that their decisions were about money and they were informed about the conversion from MUs to real money.

All games were independent, which means that in each round participants were paired with another player B. In the trust and punishment game, each round was computed as a one-shot game followed by feedback. In the cooperation game, player A played two rounds with the same players B. All games are illustrated in Figure 1.



The Trust Game consisted of two conditions: the trust condition and the non-social risk condition. Every MU retained by the players was compensated with 0.5 Swiss francs. In the trust game players A and B each had an initial amount of 10 MUs. Player A could choose an amount, in the range of 0 to 10, to transfer to player B. The transferred MUs were multiplied by three and given to player B. Player B was informed about the transfer and then decided how many MUs to transfer back to player A. She could transfer any amount within a range of 0 to 10, plus the received transfer from player A. The final payoff for player A consisted of the initial MUs minus the MUs given to player B plus the MUs received from player B. In order to compare trusting behavior with a corresponding non-social risk, we conducted another experiment in which patients and controls faced the same choices as in the trust experiment, but the decisions were not embedded in a social interaction. In this non-social risk condition, player B was replaced by a computer decision. Participants were told that they are interacting with a computer and that the computer makes its decision by randomly drawing from Player's B response distribution. Responses were uniformly distributed around the mean in the range of  $\pm 2$  SDs. Participants were informed of this calculation and the real behavioral data underlying it.

Punishment behavior was investigated using a dictator game with and without the punishment stage (Spitzer et al., 2007). In this game, Player A had an initial amount of 100 MUs, from which she could transfer an amount between 0 and 100 to player B. Player B's initial amount equaled 0 MUs. After the transfer, player A and player B received an additional endowment of 25 MUs. In the dictator condition without the punishment stage, the round finished here. In the punishment condition, player B was allowed to punish player A by deducting MUs from player A's income. Punishment was costly: if player B invested one MU from her income, this resulted in a deduction of five MUs of A's income. Player B could invest a maximum of 25 MUs

for punishment, which corresponded to a maximum deduction of 125 MUs. In this game, every MU retained was compensated with 0.13 Swiss francs.

Our Cooperation Game was a variant of the minimum effort game devised by Huyck et al. (Huyck et al., 1990). Each player interacted with 3 other players. This game was played over 10 rounds, of which 2 consecutive rounds were played with the same players. All four players simultaneously chose one of the three effort levels for each round: 5, 7.5, or 10 Swiss francs. The players who chose the minimum that was chosen by the 4 players received their chosen number as payoff. For those who chose a higher number than the minimum, the difference between the chosen option and minimum was subtracted from the minimum. The resulting payoff table of this game is shown in Tab 3 of Figure 1. In this game, every MU was rewarded with 1 Swiss franc.

## 2.3 Statistical analysis

We performed consistency analysis using Cronbach's alpha for all game conditions (trust–non-social risk, cooperation round 1–cooperation round 2, punishment condition– dictator condition) in order to quantify behavioral consistency and compare groups. Cronbach's alpha is calculated from pairwise correlations between the single-shot rounds for each group separately.

$$\alpha_g = \frac{N * \bar{r}}{1 + (N - 1) * \bar{r}} \quad (1)$$

$N$  corresponds to the number of rounds in a game,  $r$  corresponds to the averaged correlation between the rounds and  $g$  stands for the group index (BPD, MDD or HC). The Fisher-Bonett test (for small samples,  $N < 100$ ) was used to compare groups.

In order to obtain an individual measurement of consistency, we calculated the standard deviation (SD) for each participant  $j$  depending on condition (trust, non-

social risk, punishment, dictator, cooperation round 1 and cooperation round 2).  $SD_j$  measures the amount of variation (or dispersion) from the mean  $\bar{y}_j$  of a participant  $j$ . Hence, a low SD indicates that all measured values  $m$  are close to the individual mean  $\bar{y}_j$ —the participant is consistent in what she is doing. A high SD indicates that the measured values are spread out over a large range—the participant is less consistent.

$$SD_j = \sqrt{\frac{\sum_{m=1}^n (y_{jm} - \bar{y}_j)^2}{n-1}} \quad (2)$$

Furthermore, we performed analysis of variance in order to investigate mean differences in investment between the three diagnostic groups (control, MDD, BPD) depending on game.

**Trust Game:** We calculated repeated measures ANOVA with “condition” (trust, non-social risk) and “round” (rounds 1–4) as within-subject factors and “diagnosis” as the between-subject factor.

**Punishment Game:** We calculated repeated measures ANOVA with “game type” (punishment, dictator) and “round” (rounds 1–5) as within-subject factors and “diagnosis” as the between-subject factor.

**Cooperation Game:** In the cooperation game, players played two rounds as part of the same group. Therefore, investment in round 2 depended on the minimum in round 1. We calculated a multilevel regression for repeated measures (Hoffman and Rovine, 2007). This method enabled us to analyze the cooperation behavior in round 2<sub>*i*</sub> of each participant  $j$  depending on the minimum in round 1. The dependent variable was the cooperation behavior (investment) in round 2. We proposed that the minimum in round 1 predicted cooperation behavior in round 2.

The Level 1 predictor was the grand mean-centered minimum in round 1. The Level 2 predictor was diagnosis. Two dummy variables represented the control group and the MD group, whereas the BPD group was assigned the reference category.

Level 1: (3)

$$Y_{ij} = \beta_{0j} + \beta_{1j} \cdot \text{Minimum1} + \varepsilon_{ij}$$

Level 2: (4)

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \cdot \text{MD} + \gamma_{02} \cdot \text{Control} + v_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} \cdot \text{MD} + \gamma_{12} \cdot \text{Control} + v_{1j}$$

Parameter  $\gamma_{00}$  is the fixed effect of the intercept and reflects investment behavior of the BPD group in round 2 for an average minimum in round 1. Parameter  $\gamma_{01}$  and  $\gamma_{02}$  describe how the investment behavior (or  $\gamma_{00}$ ) changes for the MD and control group respectively. Parameter  $\gamma_{10}$  constitutes the fixed effect of the slope, and in this case describes the increase of investment behavior of the BPD group in round 2, depending on the minimum in round 1. Parameters  $\gamma_{11}$  and  $\gamma_{12}$  then describe how  $\gamma_{10}$  changes for the MD and control group respectively. Parameters  $v_{0j}$  and  $v_{1j}$  finally describe the random intercept and random slope for each participant.

### 3. Results

Demographic and clinical characteristics of the 17 BPD patients, 24 MDD patients, and 36 HC are presented in Table 1 and 2. HC, BPD and MDD groups were frequency matched according to education. HC and BPD were also matched according to age.

#### 3.1 Inconsistency behavior in social-decision experiments

Fisher-Bonett tests were used to compare the patient groups to the control group. Results are presented in Table 3. BPD patients significantly differed in consistency when compared to HC during the trust condition and the punishment condition ( $p \leq 0.01$ ). When compared to participants with MDD, BPD individuals exhibited a significant difference in the trust game ( $p = 0.02$ ), showing themselves to be less consistent, and but only a trend towards significance in the punishment condition ( $p = 0.07$ ). There were no differences in the non-social risk condition of the trust game and the dictator game ( $p > 0.2$ ). There was neither a significant difference in the first round of the cooperation game between BPD and HC ( $p = 0.07$ ) nor when BPD individuals were compared to MDD participants ( $p = 0.21$ ). MDD patients showed a greater consistency in the second round of the cooperation game when compared to HC ( $p = 0.048$ ) and but not compared to BPD ( $p = 0.07$ ).

The SDs were correlated with the Social support values obtained from the *Questionnaire on Social Support* for each condition. Correlation coefficients and p-values are presented in Table 4. There was a significant negative correlation between the SD in the trust condition and perceived social support ( $p = 0.04$ ), meaning that the SD decreased as perceived social support increased. Furthermore, there was a significant negative correlation between the SD in the punishment condition and perceived social support ( $p = 0.03$ ), showing that the SD in the punishment condition decreased when perceived social support increased. This effect was more pronounced for the BPD patient groups ( $p \leq 0.08$ ) than for the healthy controls ( $p \leq 0.32$ ).

### **3.2 Average investments in social-decision experiments**

Trust Game: A repeated measures ANOVA revealed no significant differences between the trust game and the non-social risk condition,  $F(1, 74) = 0.7$ ,  $p = 0.41$ ,  $\eta^2$

= 0.01 (Figure 2). Investment increased with number of rounds in the trust game,  $F(3, 6) = 11.52, p < 0.01, \eta^2 = 0.14$ , but not in the non-social risk condition,  $F(3, 6) = 0.23, p = 0.88, \eta^2 = 0.003$ . This difference was significant,  $F(3, 222) = 4.1, p = 0.007, \eta^2 = 0.05$ . Diagnostic groups did not differ in overall investment averaged across both the trust game and the non-social risk condition,  $F(2, 74) = 0.28, p = 0.76, \eta^2 = 0.01$ . None of the remaining interactions reached significance ( $p > 0.7$ )

**Punishment Game:** A repeated measures ANOVA revealed a significant difference between investment in the punishment game and investment in the dictator game,  $F(1, 74) = 31.03, p < 0.01, \eta^2 = 0.3$  (Figure 3). Investment was higher in the punishment game than in the dictator game. There was no main effect for round,  $F(4, 296) = 1.81, p = 0.13, \eta^2 = 0.02$  and no difference between the diagnostic groups,  $F(1, 74) = 0.32, p = 0.73, \eta^2 = 0.01$ . None of the interactions were significant ( $p > 0.6$ ).

**Cooperation Game:** The random intercept model without predictors (null-model) revealed a significant proportion of inter-individual variance in cooperation behaviors in round 2,  $s^2(u_{0j}) = 0.66, z = 3.19, p = 0.001$ . The intra-class correlation was  $0.19 (s^2(u_{0j}) / (s^2(u_{0j}) + s^2(\epsilon_{ij})))$ ,  $s^2(\epsilon_{ij}) = 2.9, z = 12.41, p < 0.01$ .

The results showed that the mean investment in round 2 was  $\gamma_{00} = 7.55$ ,  $F(1, 70) = 822.34, p < 0.001$ . There was no difference in cooperation behavior in the second round between MDD patients and BPD patients,  $\gamma_{01} = -0.33, F(1, 72) = 0.93, p = 0.34$ , and between HC and BPD patients,  $\gamma_{02} = -0.06, F(1, 71) = 0.04, p = 0.85$ . However, the minimum in round 1 significantly predicted cooperation behavior in round 2,  $\gamma_{10} = 0.4, F(1, 52) = 4.29, p = 0.04$ . Cooperation behavior (round 2) increased with minimum in round 1. None of the interactions were significant ( $p > 0.37$ ), indicating that the minimum in round 1 did not differentially affect the

cooperation behavior in round 2 depending on diagnosis (see supplementary material for results summary table).

## 4. Discussion

Inconsistent behavior in BPD was found in the trust and punishment game. Importantly, inconsistencies were only observed in the social risk conditions of these games. There were no mean differences in behavior between the different diagnostic groups. Furthermore, BPD, MDD, and HC participants did not differ in their dependency of payoff 2 on payoff 1 in the cooperation game. Interestingly, perceived social support, measured with the *Questionnaire on Social Support*, was strongly and negatively correlated with behavioral inconsistency in the trust and punishment conditions.

In contrast to King-Cases et al. (2008) and Unoka et al. (2009) we used one-shot trials in two of our experiments in which each interaction can be considered a first encounter with an unfamiliar partner. We were now able to show that a social context results in inconsistent social behavior in BPD patients. BPD patients were less consistent in the trust and the punishment game when compared to HC and MDD. We propose that BPD patients are not only unpredictable when it comes to social interactions; they also might show a generally less consistent behavior when it comes to first encounters. An explanation for the phenomenon will be provided in the next section.

As also described in the introduction, unpredictable and inconsistent behavior in BPD may be the result of experienced parental rejection and inconsistent treatment by caregivers during childhood (Zanarini et al., 2000). Fear of abandonment is one of the core features in BPD (Berenson et al., 2011; Rosenbach and Renneberg, 2014; Staebler et al., 2011). Rosenbach and Renneberg (2014)

recently were able to show that 'rejection sensitivity' mediates the relationship between parental rejection and BPD characteristics. This means, parental rejection is only associated with BPD symptoms in patients with high rejection sensitivity. Interestingly, perceived social support serves as protective factor, which diminishes the risk of developing BPD.

Consistent with this assumption, we were able to show that perceived social support was lower in BPD patients when compared to healthy controls. Moreover, perceived social support ratings correlated negatively with behavioral inconsistencies in the games: the lower the perceived social support scores, the higher the behavioral inconsistency. Although we cannot draw a causal conclusion from this finding, based on the attachment theory it is possible that inconsistent, neglecting or abusing treatment of caregivers during childhood leads to a disturbed ability to mentalize about others' intentions, which in turn results in the belief that others are unreliable and uncaring. This inner representation of others' intentions results in uncertain, inconsistent and unpredictable social behaviors in BPD patients and also guides inconsistent behavior in first encounters.

We did not find any mean differences in the trust and punishment game between BPD and the control groups, although a previous study by Unoka et al. (2009) reported reduced trust in BPD across 5 consecutive transactions. Furthermore, HC and MDD increased their investment across five rounds whereas BPD patients did not. In contrast to our study, participants were in each round interacting with the same trustee but did not receive feedback about the back transfer. The authors proposed that in the absence of feedback, BPD patients lose trust during sequential transactions. In the present study participants received feedback and they were in each round interacting with another trustee. We found a strong linear relationship in the social risk condition of the trust game; investment



increased with number of rounds. This finding suggests that our participants learnt from the 'trustee population' during each round. Interestingly, Franzen et al. (2011) proposed a superior 'theory of mind' in BPD because BPD patients adjusted their investment to the objective fairness of a partner in a multiround trust game, whereas healthy controls disregard the objective fairness of a partner when emotional facial expressions are present. Although we were able to control for reputation effects we can hence not fully exclude learning from feedback. It would therefore be interesting to investigate these learning effects in future studies.

Although we were able to provide preliminary evidence for inconsistent interpersonal behavior in BPD patients using a game theoretic approach, several limitations merit comment. We were not able to confirm our hypothesis for the cooperation game. BPD patients showed only a trend towards significance for a less consistent investment behavior in the first round of the cooperation game compared to HC. Furthermore, the patient groups did not differ in their dependency of payoff 2 on payoff 1 in the cooperation game. Additionally, our sample size was relatively small and we only tested female participants, which was due to the higher prevalence of BPD in women. Therefore, future studies should aim to replicate the present findings. Furthermore, MDD patients were significantly older than the BPD patients. Fett et al. (2014) investigated trust and cooperation in groups of subjects ranging from adolescence to mid-adulthood. They reported that increasing age is associated with higher trust. Sutter and Kocher (2007) examined trust and trustworthiness in different age groups and they reported a linear increase from early childhood to early adulthood. Within the different adult age groups, however, trust remains constant. We only tested adult participants. Therefore, we propose that differences in BPD and MDD patients are most likely not due to age difference. However, no study has yet

investigated behavioral consistency in social decision-making across different age group.

Future studies should involve imaging techniques in order to provide further understanding of the neural network underlying BPD-specific inconsistent behavior. Areas such as the anterior insula, the lateral orbitofrontal cortex and the right dorsolateral prefrontal cortex have been implicated as target regions of BPD (King-Casas et al., 2008; Seres et al., 2009; Spitzer et al., 2007). Furthermore, Berdahl (2010) suggested the involvement of the vmPFC in a neural network model of BPD. This area is important for decisions regarding social preferences. The games used in the present study would be well suited to investigate the neural basis underlying behavioral inconsistency in BPD.

In conclusion, the results of our study suggest that BPD patients behave inconsistently in different kinds of social interactions games even in first encounters. Furthermore, inconsistencies in social behavior were associated with a lack of perceived social support. To the best of our knowledge, this is the first study to show behavioral inconsistencies in BPD in an experimental design. Future studies should investigate the neural basis of behavioral inconsistency in BPD and whether inconsistent behavior is also evident in multi-round games.

### **Financial Disclosures**

The authors do not have conflicts of interest regarding the content of this study.

### **Acknowledgements**

The authors would like to thank Ulrich Raub and Roger Gubelmann for their support in recruiting patients for this study.

## References

- Ainsworth, M.D.S., Blehar, M.C., Waters, E., Wall, S., 1978. Patterns of attachment: a psychological study of the strange situation. *Am. Psychol.*, 44, 709–16.
- Agrawal, H.R., Gunderson, J., Holmes, B.M., Lyons-Ruth, K., 2004. Attachment studies with borderline patients: a review. *Harv. Rev. of Psychiatry* 12 (2), 94-104.
- Ball, J. S., & Links, P. S., 2009. Borderline personality disorder and childhood trauma: evidence for a causal relationship. *Curr. Psychiatry Rep.*, 11(1), 63-68.
- Beblo, T., Saavedra, A.S., Mensebach, C., Lange, W., Markowitsch, H.-J., Rau, H., Woermann, F.G., Driessen, M., 2006. Deficits in visual functions and neuropsychological inconsistency in Borderline Personality Disorder. *Psychiat. Res.* 145 (2–3), 127-135.
- Beck, A.T., Epstein, N., Brown, G., Steer, R.A., 1988. An inventory for measuring clinical anxiety: psychometric properties. *J. Consult. Clin. Psych.* 56 (6), 893-897.
- Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., Erbaugh, J., 1961. An inventory for measuring depression. *Arch. Gen. Psychiatry* 4, 561-571.
- Berdahl, C.H., 2010. A neural network model of Borderline Personality Disorder. *Neural Networks* 23 (2), 177-188.
- Berenson, K.R., Downey, G., Rafaeli, E., Coifman, K.G., Paquin, N.L., 2011. The rejection–rage contingency in borderline personality disorder. *J. Abnorm. Psychol.* 120 (3), 681.
- Bowlby, J. 1969. *Attachment and Loss: Attachment* (Vol. 1). Hogarth Press and the Institute of Psycho-Analysis, London.
- Bowlby, J., 1973. *Attachment and loss: Separation anger and anxiety* (Vol. 2). Basic Books, New York.
- Bowlby, J., 1977. The making and breaking of affectional bonds. II. Some principles of psychotherapy. The fiftieth Maudsley Lecture. *Brit. J. Psychiat.* 130 (5), 421-431.
- Bowlby, J., 1980. *Attachment and loss: Loss, sadness and depression* (Vol. 3). Basic Books, New York.
- Clarkin, J.F., Lenzenweger, M.F., Yeomans, F., Levy, K.N & Kernberg, O.F., 2007. An object relations model of borderline pathology. *J. Pers. Disord.* 21(5), 474-499.
- Doll, A., Sorg, C., Manoliu, A., Meng, C., Wöller, A., Förstl, H., Zimmer, C., Wohlschläger, A., Riedl, V., 2013. Shifted intrinsic connectivity of central executive and salience network in borderline personality disorder. *Fr. Hum. Neurosc.* 7.
- Fett, A.-K.J., Gromann, P.M., Giampietro, V., Shergill, S.S., Krabbendam, L., 2014. Default distrust? An fMRI investigation of the neural development of trust and cooperation. *Soc. Cogn. Affect. Neurosci.* 9 (4), 395-402.
- First, M.B., Gibbon, M., Spitzer, R.L., Williams, J.B.W., Benjamin, L.S., 1997. *Structured Clinical Interview for DSM-IV Axis II Personality Disorders (SCID-II)*. American Psychiatric Press, Inc., Washington, D.C.
- First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 2002. *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP)*. Biometrics Research, New York State Psychiatric Institute, New York.
- Fonagy, P., 2000. Attachment and borderline personality disorder. *J. Am. Psychoanal. Ass.*, 48(4), 1129-1146.
- Fonagy, P., Target, M., & Gergely, G., 2000. Attachment and borderline personality disorder: A theory and some evidence. *Psychiat. Clin. N. Am.*, 23(1), 103-122.
- Franzen, N., Hagenhoff, M., Baer, N., Schmidt, A., Mier, D., Sammer, G., Gallhofer, B., Kirsch, P., Lis, S., 2011. Superior ‘theory of mind’ in borderline personality disorder: An analysis of interaction behavior in a virtual trust game. *Psychiat. Res.* 187 (1–2), 224-233.
- Fydrich, T., Sommer, G., Tydecks, S., Brähler, E., 2009. Fragebogen zur sozialen Unterstützung (F-SozU): Normierung der Kurzform (K-14) Social Support Questionnaire (F-SozU): Standardization of short form (K-14). *Zeitschrift für Medizinische Psychologie* 18 (1), 43-48.

- Hoffman, L., Rovine, M., 2007. Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behav. Res. Methods* 39 (1), 101-117.
- Huyck, J.B.V., Battalio, R.C., Beil, R.O., 1990. Tacit coordination games, strategic uncertainty, and coordination failure. *Am. Econ. Rev.* 80 (1), 234-248.
- King-Casas, B., Sharp, C., Lomax-Bream, L., Lohrenz, T., Fonagy, P., Montague, P.R., 2008. The rupture and repair of cooperation in borderline personality disorder. *Science* 321 (5890), 806-810.
- Leichsenring, F., Leibing, E., Kruse, J., New, A.S., Leweke, F., 2011. Borderline personality disorder. *Lancet* 377 (9759), 74-84.
- Levy, K.N., 2005. The implications of attachment theory and research for understanding borderline personality disorder. *Dev. Psychopathol.* 17 (04), 959-986.
- Lieb, K., Zanarini, M. C., Schmahl, C., Linehan, M. M., & Bohus, M., 2004. Borderline personality disorder. *Lancet*, 364(9432), 453-461.
- Rosenbach, C., Renneberg, B., 2014. Rejection sensitivity as a mediator of the relationship between experienced rejection and borderline characteristics. *Pers. Individ. Differ.* 69, 176-181.
- Schmideberg, M., 1959. "The borderline patient," in *American Handbook of Psychiatry*. Basic Books, New York.
- Seres, I., Unoka, Z., Keri, S., 2009. The broken trust and cooperation in borderline personality disorder. *Neuroreport* 20 (4), 388-392.
- Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., Fehr, E., 2007. The neural signature of social norm compliance. *Neuron* 56 (1), 185-196.
- Staebler, K., Helbing, E., Rosenbach, C., Renneberg, B., 2011. Rejection sensitivity and borderline personality disorder. *Clin. Psychol. Psychot.* 18 (4), 275-283.
- Sutter, M., Kocher, M.G., 2007. Trust and trustworthiness across different age groups. *Game Econ. Behav.* 59 (2), 364-382.
- Unoka, Z., Seres, I., Áspán, N., Bódi, N., Kéri, S., 2009. Trust game reveals restricted interpersonal transactions in patients with borderline personality disorder. *J. Pers. Disord.* 23 (4), 399-409.
- Zanarini, M.C., 2003. Zanarini Rating Scale for Borderline Personality Disorder (ZAN-BPD): a continuous measure of DSM-IV borderline psychopathology. *J. Pers. Disord.* 17 (3), 233-242.
- Zanarini, M.C., Ruser, T.F., Frankenburg, F.R., Hennen, J., Gunderson, J.G., 2000. Risk factors associated with the dissociative experiences of borderline patients. *J. Nerv. Ment. Dis.* 188 (1), 26-30.

Table 1: Demographical data

	<i>HC</i>	<i>BPD</i>	<i>MDD</i>	<i>Statistics</i>
<b>Age</b>	36.97 (11.95)	30.53 (11.06)	43.63 (14.02)	$F(2, 74) = 5.6, p = 0.005^a$
<b>Education</b>	5 (1.87)	4 (1.12)	4.21 (1.64)	n.s.
<b>BDI</b>	1.5 (2.51)	19.88 (14.11)	26.5 (10.43)	$F(2, 74) = 62.19, p < 0.001^b$
<b>BAI</b>	2.03 (2.08)	14.18 (9.99)	20.33 (12.69)	$F(2, 74) = 35.02, p < 0.001^c$
<b>Zan-BPD</b>	0.06 (0.23)	10.82 (6)	6.46 (19.94)	$F(2, 73) = 63.03, p < 0.001^d$
<b>GAF</b>	88.61 (2.46)	57.53 (12.86)	55.08 (12.63)	$F(2, 74) = 115.22, p < 0.001^e$
<b>BMI</b>	22.74 (3.51)	26.21 (6.75)	25.71 (7.91)	n.s.
<b>Social support</b>	4.48 (0.47)	3.63 (0.84)	3.89 (0.65)	$F(2, 74) = 12.84, p < 0.001^f$
<b>Compensation Money</b>	46.16 (9.07)	45.44 (7.8)	46.46 (8.1)	n.s.

HC, healthy controls; BPD, borderline personality disorder; MDD, major depressive disorder. BDI, Beck Depression Inventory; BAI, Beck Anxiety Inventory; GAF, Global Assessment of Functioning Scale; Social support, Questionnaire on Social Support; ZAN-BPD, Zanarini Rating Scale for BPD. Standard deviations are indicated in brackets.

<sup>a</sup>Posthoc testing: MDD patients were significantly older than BPD patients ( $t(39) = 3.21, p = 0.003$ ). There was no significant difference in age between HC and MDD or HC and BPD.

<sup>b</sup>Posthoc testing: In comparison to HC, scores for BDI were higher in MDD ( $t(24.79) = -11.52, p < 0.001$ ) and in BPD ( $t(16.48) = -5.33, p < 0.001$ ). There was no significant difference in BDI scores between BPD and MDD.

<sup>c</sup>Posthoc testing: In comparison to HC, scores for BAI were higher in MDD ( $t(23.82) = -7.01, p < 0.001$ ) and in BPD ( $t(16.67) = -5.00, p < 0.001$ ). There was no significant difference in BAI scores between BPD and MDD.

<sup>d</sup>Posthoc testing: BPD patients had a higher overall ZAN-BPD score than HC ( $t(16.02) = -7.39, p < 0.001$ ) and MDD ( $t(22.19) = -5.27, p < 0.001$ ). MDD patients had a higher overall ZAN-BPD score than HC ( $t(22.16) = -3.71, p = 0.001$ ).

<sup>e</sup>Posthoc testing: Although GAF was higher for the subjects with MDD than for the healthy controls ( $t(24.17) = 12.85, p < 0.001$ ), there was no difference in GAF scores between BPD and MDD. The GAF score of 88.61 points for the healthy group indicated absent or minimal symptoms, good functioning in all areas, and interest and involvement in a wide range of activities. The score of 55.08 obtained by the MDD patients indicated the presence of moderate difficulties in social or occupational functioning.

<sup>f</sup>Posthoc testing: Social support as measured by the *Questionnaire on Social Support* was higher in HC than in BPD ( $t(20.85) = -3.89, p = 0.001$ ) or MDD ( $t(58) = 3.81, p < 0.001$ ). There was, however, no difference in social support between BPD and MDD.

Table 2: Comorbidities in Borderline personality disorder (BPD), major depressive disorder (MDD) and healthy controls (HC).

	<i><b>HC</b></i>	<i><b>in %</b></i>	<i><b>BPD</b></i>	<i><b>in %</b></i>	<i><b>MDD</b></i>	<i><b>in %</b></i>
<b>Substance Abuse</b>	0	0	3	18	1	4.2
<b>Affective Disorder</b>	0	0	7	41	NA	NA
<b>Anxiety Disorder</b>	0	0	10	59	12	50
<b>Eating Disorder</b>	0	0	9	53	6	25
<b>Smokers</b>	10	28	13	76	13	54

HC, healthy controls; BPD, borderline personality disorder; MDD, major depressive disorder.

Table 3: Results of the behavioral consistency analysis comparing Borderline personality disorder (BPD), major depressive disorder (MDD) and healthy controls (HC): BPD patients significantly differed in behavioral consistency from HC in the punishment condition of the punishment game and in the social-risk condition of the trust game.

	<i>Group</i>	<i>Cronbach's alpha</i>	<i>Sample size</i>	<i>No. of items</i>	<i>Z statistics</i> <i>BPD &amp; MDD vs. control</i>	<i>p</i>	<i>Z statistics</i> <i>BPD vs. MDD</i>	<i>p</i>
<b>Punishment Condition</b>	HC	0.886	36	5				
	BPD	0.66	17	5	-2.23	<b>0.01**</b>		
	MDD	0.847	24	5	-0.68	0.25	-1.51	0.066 <sup>a</sup>
<b>Dictator Condition</b>	HC	0.914	36	5				
	BPD	0.875	17	5	-0.76	0.22		
	MDD	0.91	24	5	-0.11	0.46	-0.62	0.27
<b>Trust Condition</b>	HC	0.79	36	4				
	BPD	0.326	17	4	-2.38	<b>0.009**</b>		
	MDD	0.79	24	4	0	0.5	-2.13	<b>0.02*</b>
<b>Non-Social-Risk Condition</b>	HC	0.737	36	4				
	BPD	0.739	17	4	0.02	.51		
	MDD	0.616	24	4	-0.85	.2	-0.71	.24
<b>Cooperation Rounds 1</b>	HC	0.675	36	5				
	BPD	0.34	17	5	-1.45	.07 <sup>a</sup>		
	MDD	0.571	24	5	-0.64	.26	-0.81	0.21

<b>Cooperation Rounds 2</b>	HC	0.397	36	5				
	BPD	0.361	17	5	-0.12	.45		
	MDD	0.706	24	5	-1.66	<b>.048*</b>	-1.47	.07 <sup>a</sup>

HC, healthy controls; BPD, borderline personality disorder; MDD, major depressive disorder

Table 4: Results of Spearman's rank order correlation between standard deviation of decision-making (SD) and social support.

	<i><b>SD trust</b></i>	<i><b>SD nonsocial risk</b></i>	<i><b>SD punishment</b></i>	<i><b>SD dictator</b></i>	<i><b>SD Coop 1</b></i>	<i><b>SD Coop 2</b></i>
<b>Social support<sup>1</sup></b>	<b>-0.24*</b>	-0.19	<b>-0.25*</b>	-0.07	-0.01	-0.11
<b>p</b>	<b>0.04</b>	0.09	<b>0.03</b>	0.54	0.94	0.33

<sup>1</sup>Social support, Questionnaire on social support; SD, standard deviation

\*p<.05



## Figure legend

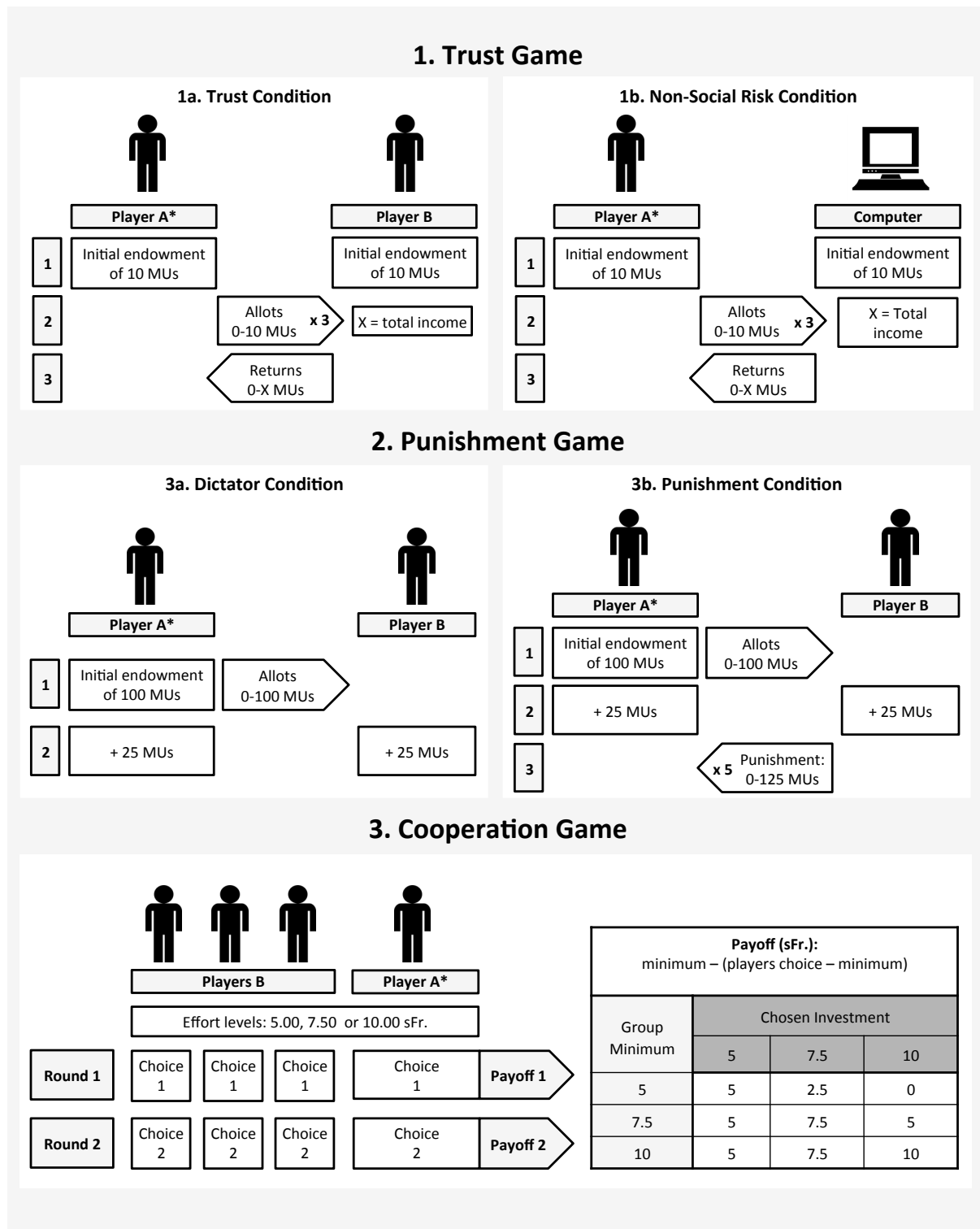
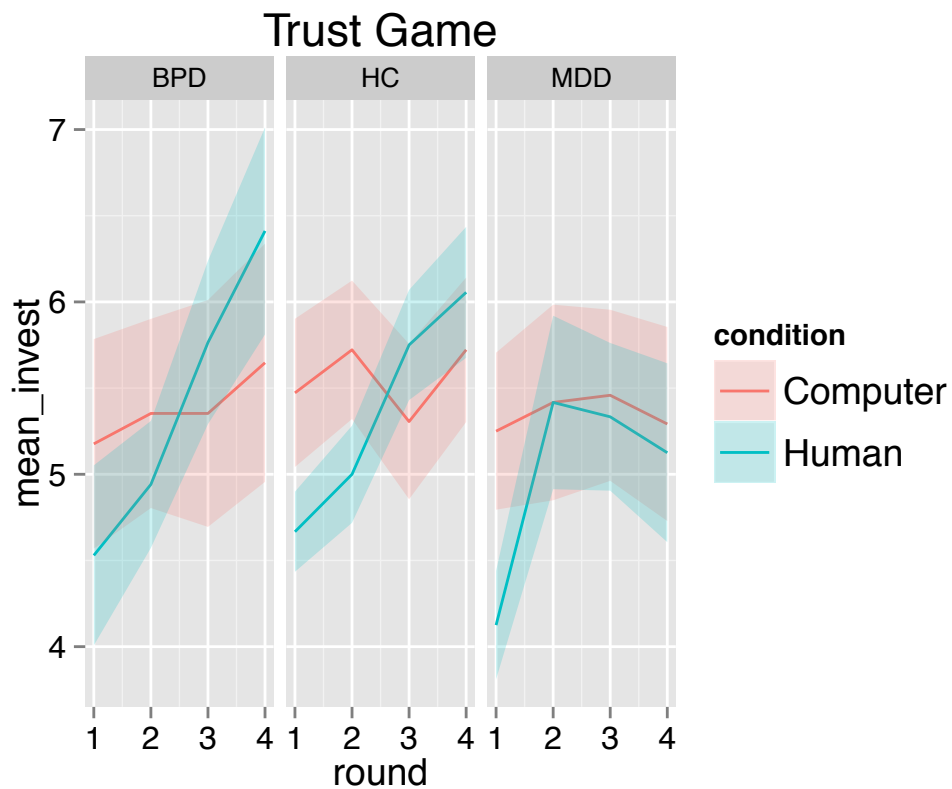
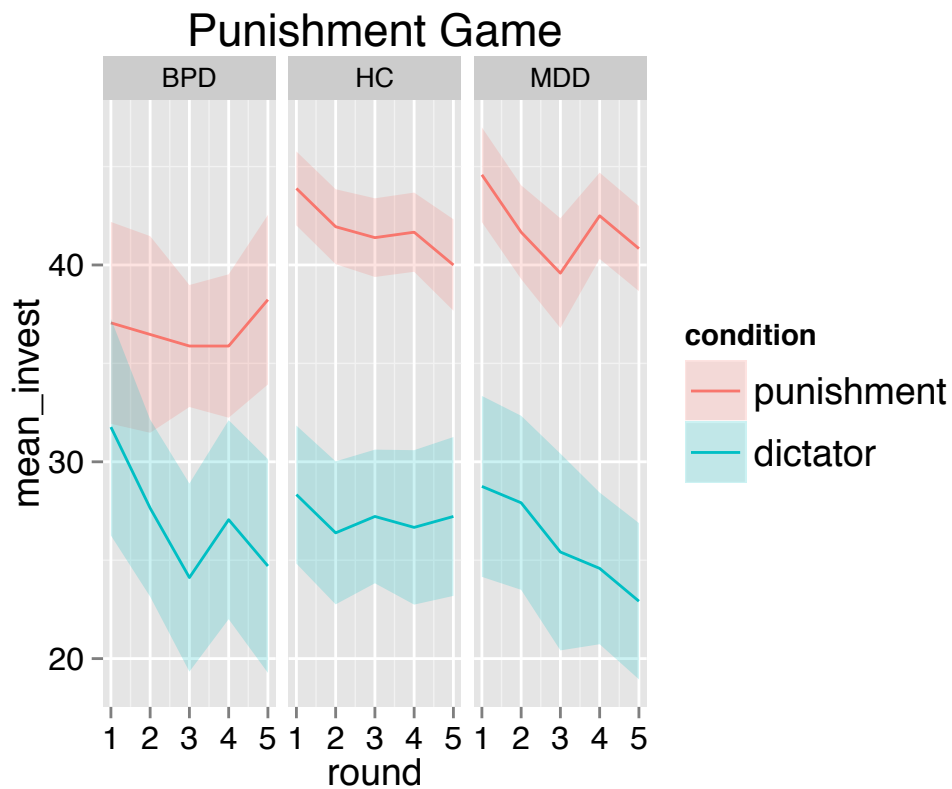


Figure 1. Graphical illustration of the social games: 1. Trust Game, 2. Punishment Game, 3. Cooperation Game.



*Figure 2.* Investment in the trust game depending on diagnostic group and condition. Analysis of variance revealed no significant difference between groups. Investment increased in the social risk condition with increasing number of rounds.



*Figure 3.* Investment in the punishment game depending on diagnostic group (Borderline personality disorder (BPD), major depression (MDD) and healthy controls (HC)) and condition: There was no mean difference between the diagnostic groups. BPD patients, however, showed a greater behavioral inconsistency, which can also be seen in the shading of the standard error in the social risk condition (punishment).